Part III. Final Project

A. Introduction and Background

The origin of the Olympics can be traced back to ancient Greece. At that time, the reasons for hosting sporting events were not only for athletic competitions between Greek city-states but also for religious celebrations and offerings to the gods. Since the ancient Olympic stadium's ruins are on Mount Olympus, the Olympics were thus named.

It wasn't until 1894 that the Frenchman Pierre de Coubertin introduced the model of the Olympics that we know today. (What Is the Origin of the Olympic Games? (n.d.). International Olympics Committee.) In addition to maintaining the tradition of holding the event every four years, as in ancient Greece, he expanded its scale to a global level. The location of the Olympic Games is now rotated around the world, which helps promote it among different countries.

As the number of Olympic events and the scale of the Games expand each year, the number of viewers watching the broadcasts has skyrocketed. According to data provided by NBCUniversal, which broadcasted the Olympics this year, the 2024 Paris Olympics attracted 30.6 million viewers, an 82% increase compared to the 2020 Tokyo Olympics (16.9 million). (NBCUniversal's Presentation of Spectacular Paris Olympics Dominates Media Landscape Across All Platforms. (n.d.). NBCUniversal.) According to the official report from the International Olympic Committee (IOC), the viewership of the 2024 Paris Olympics on television and streaming platforms increased by 15.8% compared to the 2020 Tokyo Olympics. (*IOC President Praises Broadcast Operations as Paris 2024 Reaches Record Audiences*. (n.d.). International Olympics Committee.)

The immense attention and large viewership of the Olympics generate substantial revenue. According to estimates from Sports Pro Media, this year's Olympics brought in $12 billion for Paris, with a significant portion coming from sponsorships and endorsements. Olympic athletes also need to sign endorsement contracts with these sponsors, as government subsidies and medal bonuses from many countries are often insufficient to cover the high costs of training and competing year after year. For example, track and field athlete Simone Biles earned $8.5 million in endorsement deals in 2023, while swimmers Katie Ledecky and Caeleb Dressel each signed contracts worth millions of dollars. (Mcmillen, J. 2024, July 24.)

From these examples, it's clear that signing endorsement contracts with already famous, medal-winning athletes can be extremely costly, beyond what most companies can afford.

However, by targeting athletes who are not initially favored by the public but actually have a strong chance of winning medals, companies can secure endorsement rights at a relatively lower price. If these athletes go on to win medals and gain fame, the return on investment would be extraordinarily high.

The goal of this report is to analyze Olympic data to identify athletes with a high probability of winning medals and use these insights for marketing decisions. Partnering with potential winners early can increase brand exposure when they succeed.

Inspired by Samsung Taiwan's partnership with badminton duo Lee Yang and Wang Chi-Lin, who weren't initially predicted to win gold but did, boosting Samsung's visibility, we aim to use data analysis to find athletes with similar potential.

We will primarily use data from the Kaggle dataset 120 Years of Olympic History: Athletes and Results, focusing on variables like height, weight, age, nationality, and results to predict future medalists.

B. Datasets

Introduction:
This is a historical dataset on the modern Olympic Games collected by Randi H. Griffin, a Data Scientist at Boston Consulting Group. She scraped the data from sports references, and the original dataset was created and maintained by independent enthusiasts of Olympic history, without funding from an official entity. The purpose of this data collection was to provide a detailed Olympic database, as the International Olympic Committee (IOC) did not maintain one.

Timeline:
The dataset covers the Olympic Games from the Athens Summer Olympics in 1896 to the Rio Summer Olympics in 2016.

Dataset Description:
The dataset contains 271,116 rows and 15 columns. Each row corresponds to an individual athlete competing in a specific Olympic event. Key fields include the athlete's name, sex, medal status, and more. The dataset includes two fields related to regions: "Team" and "City." "Team" refers to the country that the athlete represents, and "City" refers to the host city of the event.

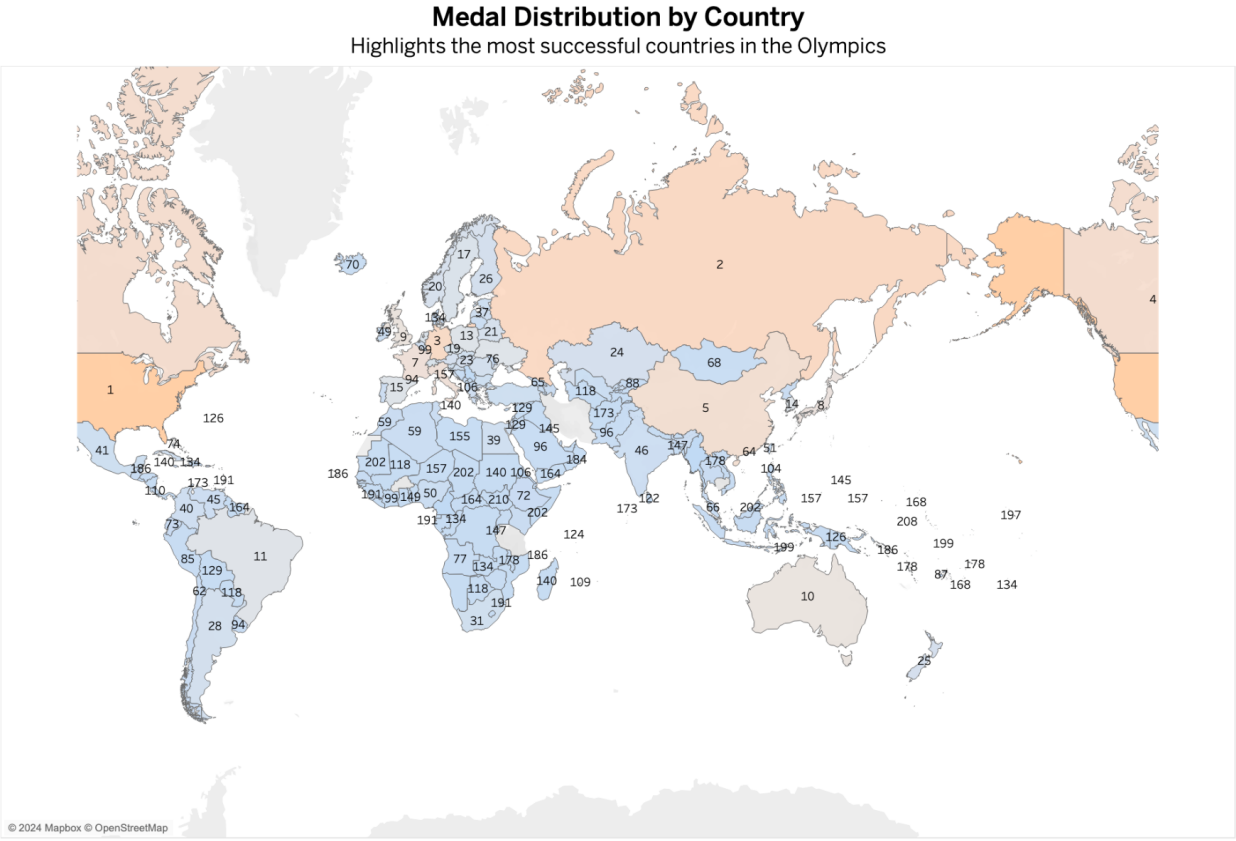
Introduction:
This dataset contains information on the GDP and population of 177 countries for the year 2022. The data was sourced from the World Bank (World Development Indicators, July 25, 2023) and the United Nations (World Population Prospects 2022).

Timeline:
The dataset reflects economic indicators from the year 2022.

Dataset Description:
The dataset consists of 177 rows and 7 columns, representing key economic data for different countries during the fiscal year 2022. These columns include each country's Gross Domestic Product (GDP), GDP growth rate, population, GDP per capita, and share of the world's total GDP. Each row corresponds to a specific country, providing a detailed look at their economic performance during this period.

C. Data story

1. Medal distribution by country



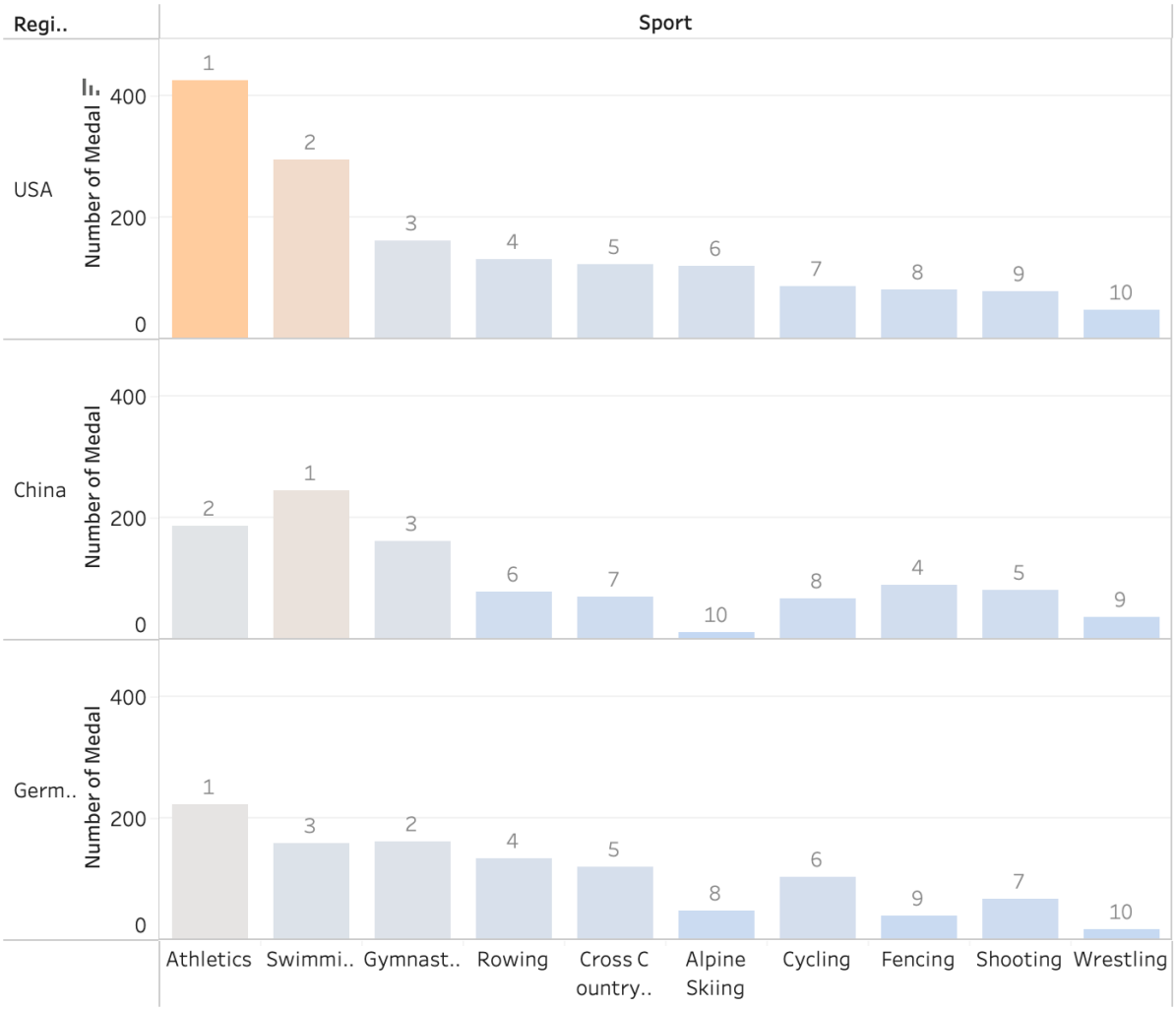
This chart displays the global ranking of Olympic medals by country. The numbers on the map represent the medal rankings of each country, with higher-ranking countries having won more medals. Countries in orange with more medals, while countries in blue with fewer medals. It can also be seen that the top-ranking countries are usually those with stronger economies. This may be because countries with better economic capabilities are more able to invest in training athletes, and athletes in these countries tend to have more promising development prospects. While less developed regions, particularly in Africa and parts of Asia, tend to win fewer medals.

Our goal is to select the athletes we want to invest in, so the first step is to narrow down the selection by choosing the countries where we want to invest in athletes. This way, we can better expect the athletes to receive quality training, ensuring a higher likelihood of success. Therefore, we select investment targets from the top five countries, which include the United States, Russia, Germany, Canada, and China. Also, to expand our market and showcase our brand to the world, we decided to invest in countries from different continents. Among the top five, although Russia ranks high, due to the current geopolitical situation and its ban from participating in the Olympics, we consider the investment risk too high. Finally, we chose Germany as the

representative for Europe. Ultimately, we decided to invest in the **United States** (North America), **Germany** (Europe), and **China** (Asia).

Chart Creation Process:
The chart was created by placing Longitude (generated) in the columns and Latitude (generated) in the rows to display countries accurately on the map. CNT(Medal) controls both color and size, with dark orange for more medals and light blue for fewer. Labels show each country's ranking and NOC and ATTR(Region) provide additional details. Filters were applied to display data from 2006-2016, with invalid coordinates excluded for accuracy.

2. Top 10 Sports by Number of Medals



Next, we want to identify which sports to invest in to achieve the highest brand exposure. We analyzed the top 10 sports with the most medals for each country. From the chart, we can see that each country excels in different sports. While the top sports may overlap, their rankings vary. This is quite evident, as body types and popular trends differ across countries. For instance, Asians tend to favor fast and agile sports, such as table tennis and badminton, whereas people from Europe and the Americas, who are generally taller and stronger, often prefer strength and endurance-based sports like basketball and football.

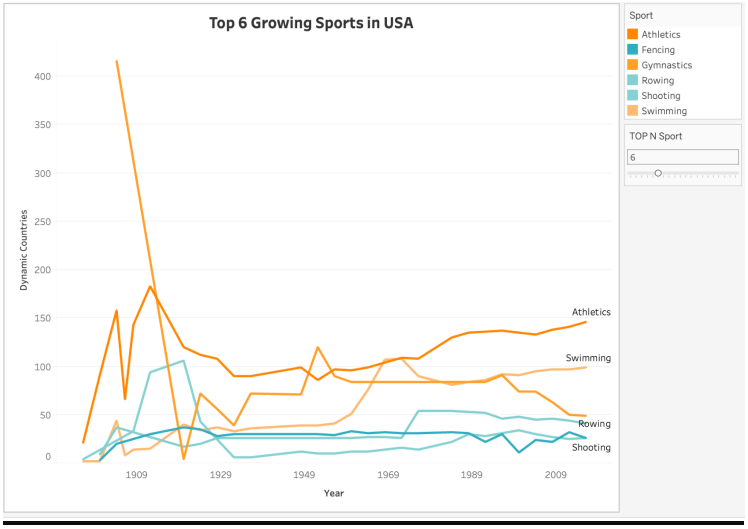
Typically, the sports in which a country wins more medals indicate a higher number of participants or more frequent competitions. This also leads to increased broadcast coverage, resulting in a larger audience for that sport. Additionally, viewers tend to prefer watching sports in which their country is more likely to win. In such cases, investing in those sports would achieve better promotional outcomes for our brand.

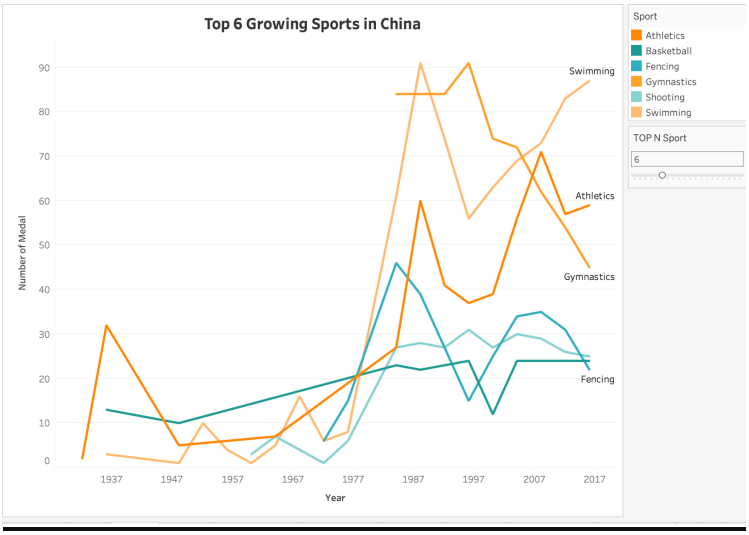
After analyzing the USA, Germany, and China, we made the following selections:

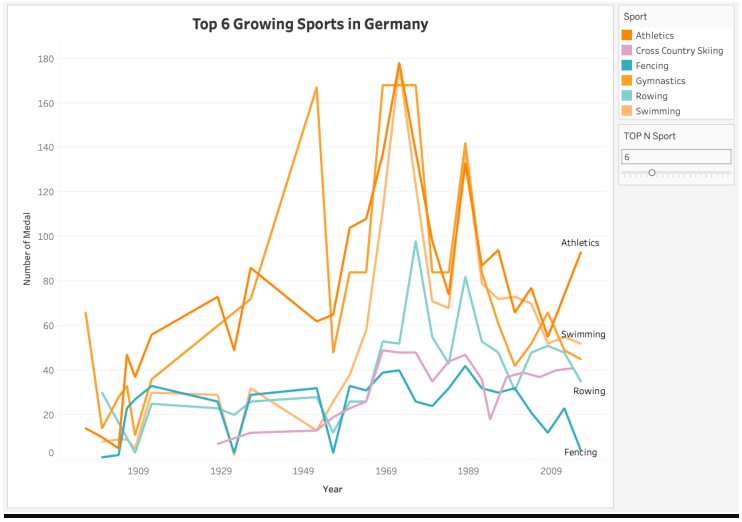
- **USA**: Athletics (as athletics leads by a large margin in the USA's Olympic medal count)
- **China**: Swimming (swimming ranks the highest in terms of medal count in China)
- **Germany**: Gymnastics (although athletics ranks first in Germany, we want to select different sports for each country)

Chart Creation Process:
To create the charts, Sport was placed in the columns to display each sport, while Region and CNT(Medal) were placed in the rows to show the selected countries and their corresponding medal counts. The chart type was set to bar, with CNT(Medal) used to control both color and size, where darker orange indicates more medals and lighter blue indicates fewer medals. Labels showing the exact number of medals were added to the bars. Filters were applied to show data from 2006 to 2016, the top 10 sports by number of medals, and data specific to the regions of the USA, Germany, and China.

3. Top 6 Growing Sports







After selecting the sports, we conducted further analysis to ensure these choices were correct by examining the growth trends of different sports in each country. The focus was on identifying sports with significant future potential and a growing audience. Through this analysis, we were able to confirm whether our investments aligned with the sports that people are increasingly interested in watching.

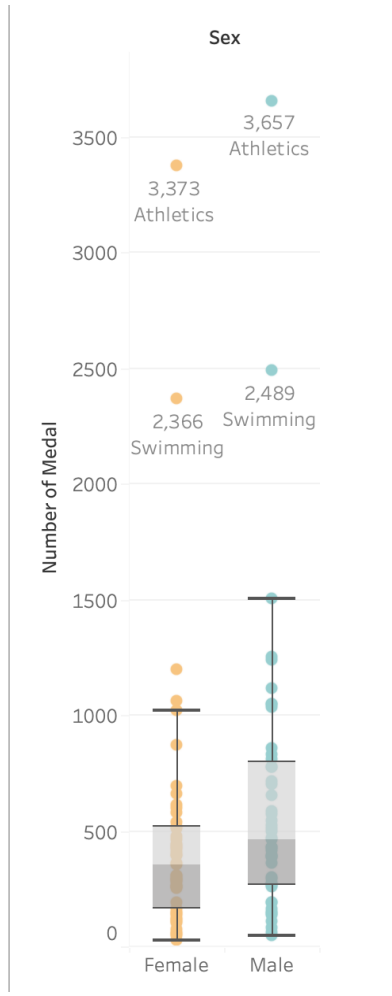
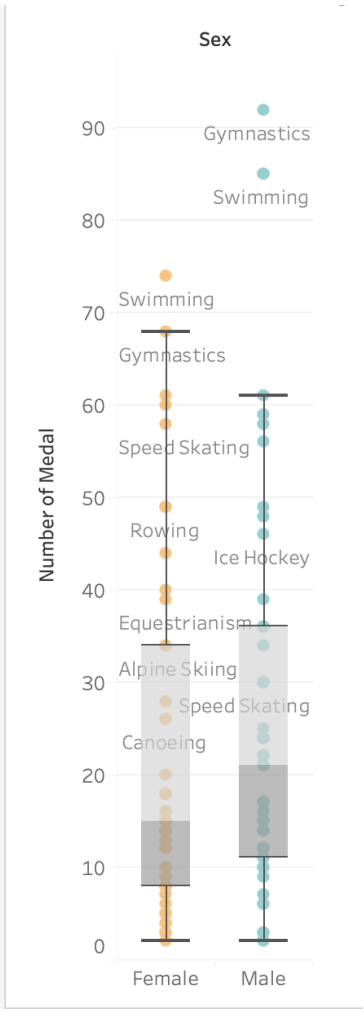
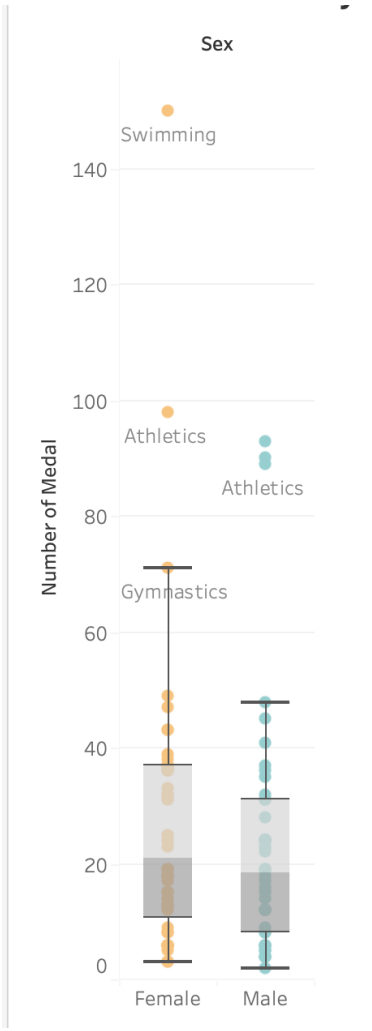
USA:
In the USA, we chose athletics as the investment focus. The data shows that athletics is one of the fastest-growing sports, with both strong Olympic performance and a rising number of participants. This trend suggests that athletics will continue to attract more viewers and media attention in future international competitions. Therefore, investing in athletics not only ensures brand visibility but also capitalizes on the sport's growth potential.

China:
In China, we chose swimming as our investment. Swimming is the fastest-growing sport in China, much like athletics in the USA. Chinese swimmers have shown excellent performance, and the sport has a solid audience base with a growing number of participants. This means investing in swimming in China will ensure continuous brand exposure, especially during upcoming Olympic events.

Germany:
For Germany, we selected gymnastics as the investment target. Although athletics and swimming also perform well in Germany, given that the USA and China have already chosen them and have a much stronger presence in those sports, we decided to invest in gymnastics. This choice allows us to further expand our market, particularly in Europe.

Chart Creation Process:
First, we placed YEAR(Year) in the Columns to display the yearly trends for each sport and added CNT(Medal) or AGG(Dynamic Countries) in the Rows to show the number of medals won or the number of participating countries. Next, we selected Line as the mark type and added Sport to Color to assign different colors to each sport. To highlight the top three growing sports, we marked the lines for these sports in orange to make them visually distinct. Then, we added a Region to the Filter to focus on the selected countries: the USA, China, and Germany. Finally, we placed Sport in Label to tag the name of each sport, allowing readers to quickly understand which trend line corresponds to each sport.

4. Medal Distribution by Gender

USA	Germany	China
		

After selecting the sports to invest in, we conducted a deeper analysis to understand the gender distribution in each sport, which helped us decide whether to invest in male or female competitions. This gender analysis is crucial for fine-tuning our investment strategy, as performance differences between men and women in certain sports can be significant. These differences not only impact the number of medals won but also influence the audience and media coverage of those events. With this in mind, we created visualizations for the USA, Germany, and China to analyze medal distribution by gender and ensure we target the gender group with the strongest advantage.

USA:

In the USA, we focused on **men's athletics**. The boxplot shows the medal distribution across all the Olympic games by gender. That is, each dot on the chart shows the number of medals in different types of sports.

Therefore, we found male athletes have a higher total count of medals compared to female athletes. Based on this analysis, we decided to invest in USA men's athletics.

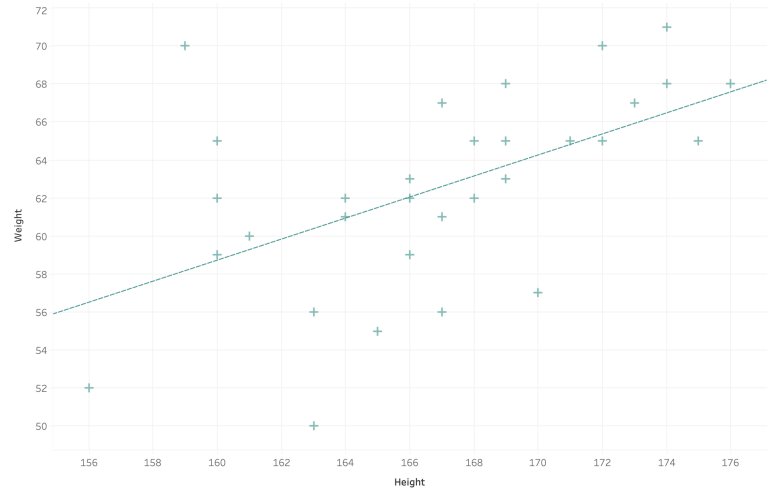
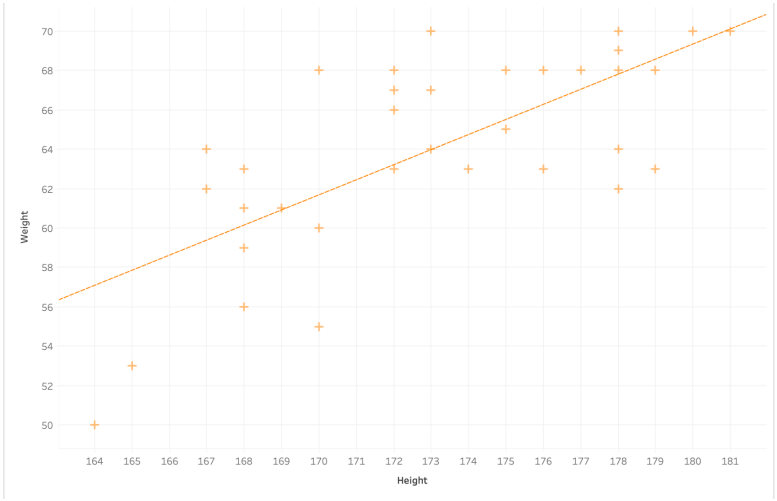
Germany:
For Germany, we analyzed **men's gymnastics**. The analysis clearly shows that male gymnasts outperform their female counterparts. The box plot highlights that men have won significantly more medals than women in gymnastics. With this insight, we chose to invest in men's gymnastics in Germany.

China:
In China, we focused on **women's swimming**. The analysis shows that female swimmers in China have won significantly more medals than their male counterparts, compared to other countries where men are more dominant. This indicates that China's female swimmers are highly competitive at the Olympic level. Based on these findings, we decided to invest in women's swimming in China.

Chart Creation Process:

First, Sex was placed in the columns to differentiate between genders, and CNT(Medal) was placed in the rows to show the number of medals. The mark type was set to a Circle, displaying the number of medals won by each gender in various sports, while the box plot indicates the data distribution range. Then, Sex was added to the color field to distinguish male and female, and Sport was added to the label field to show specific sports and their corresponding medal counts. Finally, Region and Year filters were applied to focus on the USA, Germany, and China. These steps allow for the visualization and analysis of the gender distribution in each sport

5. Weight and Height Distribution of Medalists

USA	
Germany	
China	

After identifying the specific sports and genders for investment, we analyzed the body types of athletes of the fields that we selected. The scatter plots show the relationship between height and weight for medalists in each chosen category. By examining the trends, we can find the ideal weight-to-height ratio for future talent. The trend lines in each plot give us insight into the typical body composition of successful athletes. The final weight-to-height ratios are:

USA – Men's Athletic: Weight = 1.4 * Height - 177

From the scatter plot of USA men's athletics, we can see a strong correlation between height and weight. The data points are densely distributed along the trend line, indicating a high level of consistency between changes in height and weight. The steep slope of the trend line suggests that as height increases, weight also rises significantly. This reveals a strong linear relationship between height and weight in this sport. Overall, the tightly clustered data points show that the athletes in this category have relatively similar body types, with a high degree of correlation.

Germany – Men's Gymnastics: Weight = 0.6 * Height - 30

In contrast, the scatter plot for German men's gymnastics shows a weaker correlation between height and weight. The data points are more scattered across the chart, and the trend line has a gentler slope, indicating that increases in height have a smaller impact on weight. This means that there is little difference in weight between shorter and taller athletes. The wider spread of data points also reflects greater variability in body types among the athletes, suggesting a lower degree of correlation in this sport.

China – Women's Swimming: Weight = 0.8 * Height - 68

The scatter plot for Chinese women's swimming shows a moderate correlation between height and weight. While the data points are generally aligned with the trend line, they are more dispersed compared to the USA Athletics data. The trend line's moderate slope indicates that while weight increases with height, the relationship is not as strong as in USA athletics. The data points in this chart are distributed between the tightly clustered points seen in USA athletics and the more scattered points in German gymnastics, suggesting a mix of consistency and variability in the athletes' body types.

Chart Creation Process:

To create these scatter plots, Height was placed in the columns and Weight in the rows to plot their relationship. The mark type was set to Automatic, and AVG(Age) was added to the detail field for extra context. Filters were applied to show data based on specific conditions, such as sex (male or female), region (USA, Germany, China), and sport (athletics, gymnastics, swimming). Trend lines were added to each chart to visualize the weight-to-height ratio, helping us identify the ideal body type for athletes in each categor

6. Medalists with the Ratio of Ideal Weight and Actual Weight

Madalist with Ratio of Ideal Weight and Actual Weight

Name	Weight	Height	Ideal w..	Ratio
Curtis Johnson	81.0	184.00	80.6	1.00
John Joseph "Johnny" Kel..	57.0	167.00	56.8	1.00
Thomas Francis "Tom" Fa..	64.0	172.00	63.8	1.00
Jason Pyrah	64.0	172.00	63.8	1.00
Henry Louis Scott	64.0	172.00	63.8	1.00
Helmut "Henry" Laskau	64.0	172.00	63.8	1.00
Albert Richard "Whitey" ..	64.0	172.00	63.8	1.00
Gerrard Eugene "Gene" C..	78.0	182.00	77.8	1.00
Thomas Martin "Tom" Jo..	61.0	170.00	61.0	1.00
Thomas Hugh Lilley	61.0	170.00	61.0	1.00
Thomas Charles "Tom" Ot..	61.0	170.00	61.0	1.00
Ricky Robertson	75.0	180.00	75.0	1.00
Peter Clarence Gerhardt	68.0	175.00	68.0	1.00
Paul Henry Pilgrim	75.0	180.00	75.0	1.00
Otto Kenneth Anderson	68.0	175.00	68.0	1.00
Millard Frank Hampton, Jr.	75.0	180.00	75.0	1.00
Martin William Hawkins (..	68.0	175.00	68.0	1.00
LaMark Carter	75.0	180.00	75.0	1.00
Kory Merrill Tarpenning	75.0	180.00	75.0	1.00
Kenneth Harry "Ken" Cas..	68.0	175.00	68.0	1.00
Joseph John Erxleben	61.0	170.00	61.0	1.00
Joie William Ray	54.0	165.00	54.0	1.00
James Sherman Jett	75.0	180.00	75.0	1.00
James Paul Sullivan	68.0	175.00	68.0	1.00
Horatio May Fitch	68.0	175.00	68.0	1.00
Gilbert Ronald "Ron" Larr..	54.0	165.00	54.0	1.00
Gerald Howard "Gerry" A..	75.0	180.00	75.0	1.00
George Sidney Simpson	75.0	180.00	75.0	1.00
Frederick Vaughn "Fred" ..	68.0	175.00	68.0	1.00

We ultimately decided to focus on **USA men's athletics** because the 2028 Olympics will be held in Los Angeles. Therefore, we only conducted an analysis of USA men's athletics athletes, calculating the ratio between their ideal weight (based on the trend line formula) and actual weight. A ratio of 1 indicates that the athlete's actual weight perfectly matches their ideal weight, making them the best candidates. This table helps us identify the athletes whose body composition best aligns with our ideal standards. The closer the ratio is to 1, the closer the athlete is to the ideal body type.

Chart Creation Process:

First, Measure Names were placed in the columns and Name in the rows to display each athlete's data. Then, Weight, Height, Ideal Weight, and Ratio were placed in Measure Values to show the athletes' actual weight, height, ideal weight, and the ratio between them. Filters for Sex, Sport, and Region were

applied to display data for USA male track and field athletes. These steps allow us to easily identify the athletes with a ratio of 1, who are the best fit for our ideal criteria.

Summary and Conclusions

This report provides a comprehensive analysis of Olympic data to identify potential athletes for investment, based on their likelihood of winning medals. By strategically partnering with athletes before their success, brands can increase their visibility, as demonstrated in successful partnerships like Samsung Taiwan's collaboration with Olympic badminton champions Lee Yang and Wang Chi-Lin.

Through data analysis, we explored various factors influencing athletes' success, including their nationality, sport, and physical attributes. The report highlights the importance of targeting top-performing countries, such as the United States, Germany, and China, to ensure investment in athletes who have a high probability of success. Furthermore, by focusing on popular sports in these countries—such as athletics, gymnastics, and swimming—we can maximize brand exposure due to the higher number of participants, broadcast frequency, and viewer interest in these events. Gender and body composition also play a critical role in identifying ideal athletes for investment. We can use the body shape standards obtained from past data to select the most suitable endorsement candidates with the greatest chance of earning high returns.

Finally, we decided to narrow our focus to men's athletics in the United States because the next Olympics in 2028 will be held in Los Angeles, USA. As a result, the proportion of American audiences will significantly increase. At this time, the attention on American athletes will increase, and if we invest in them, it will boost the exposure. Additionally, we found that athletes tend to perform better when competing in their home country, which suggests an increased likelihood of the U.S. winning, thereby enhancing the positive perception of our brand.

We plan to invest in either Grant Holloway or Kenny Bednarek. Grant is currently 26 years old, specializing in short-distance hurdles, standing at 188 cm and weighing 86 kg. He won one silver and one gold medal in the 2020 and 2024 Olympics respectively. Kenny, on the other hand, is 25 years old and specializes in the 200-meter race. He is also 188 cm tall and weighs 83 kg. He won two silver medals in the Olympics. Both athletes possess the ideal body proportions we have defined, and they are both highly skilled. We believe that they have the potential to be perfect brand ambassadors, bringing significant benefits to the brand.

Reference

What Is the Origin of the Olympic Games? (n.d.). International Olympics Committee.

https://olympics.com/ioc/faq/history-and-origin-of-the-games/what-is-the-origin-of-the-olympic-games

NBCUniversal's Presentation of Spectacular Paris Olympics Dominates Media Landscape Across All Platforms. (n.d.). NBCUniversal. https://www.nbcuniversal.com/article/nbcuniversals-presentation-spectacular-paris-olympics-dominates-media-landscape-across-all-platforms

IOC President Praises Broadcast Operations as Paris 2024 Reaches Record Audiences. (n.d.). International Olympics Committee. https://olympics.com/ioc/news/ioc-president-praises-broadcast-operations-as-paris-2024-reaches-record-audiences

Mcmillen, J. (2024, July 24). *Finding Gold When An Endorsement Goes Off Course: The Olympics Edition*. Forbes. https://www.forbes.com/sites/jennmcmillen/2024/07/24/finding-gold-when-an-endorsement-goes-off-course-the-olympics-edition/

Olympics 2024: Meet Team USA
https://sportsdata.usatoday.com/olympics/meet-team-usa/paris-2024

Dataset source 1: 120 years of Olympic history: athletes and results
https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results

Original data source: Sports Reference
https://www.sports-reference.com/

Dataset source 2: Worldometers - GDP by Country
https://www.worldometers.info/gdp/gdp-by-country/

Original data source: World Bank (World Development Indicators, July 25, 2023); United Nations (World Population Prospects 2022).

Contribution

Name	Contribution
Chang, Wei-Ting	Designing Introduction, Background, Objectives and Goals, Data story
Ho, Yu-Yang	Organizing and describing datasets, presentation
Huang, Tzu-Yun	Creating visualization plan, Operating Tableau
Lu, Tsung-Yu	Introduction, Background, Organizing and describing datasets
Wang, Shih-Yu	Searching Datasets, Operating Tableau, Data story, Conclusion